

Direct identification of fault estimation filter for sensor faults ^{*}

Yiming Wan ^{*} Tamas Keviczky ^{*} Michel Verhaegen ^{*}

^{*} Delft University of Technology, 2628CD, Delft, The Netherlands
(e-mail: y.wan@tudelft.nl; t.keviczky@tudelft.nl;
m.verhaegen@tudelft.nl).

Abstract: We propose a systematic method to directly identify a sensor fault estimation filter from plant input/output data collected under fault-free condition. This problem is challenging, especially when omitting the step of building an explicit state-space plant model in data-driven design, because the inverse of the underlying plant dynamics is required and needs to be stable. We show that it is possible to address this problem by relying on a system-inversion-based fault estimation filter that is parameterized using identified Markov parameters. Our novel data-driven approach improves estimation performance by avoiding the propagation of model reduction errors originating from identification of the state-space plant model into the designed filter. Furthermore, it allows additional design freedom to stabilize the obtained filter under the same stabilizability condition as the existing model-based system inversion. This crucial property enables its application to sensor faults in unstable plants, where existing data-driven filter designs could not be applied so far due to the lack of such stability guarantees (even after stabilizing the closed-loop system). A numerical simulation example of sensor faults in an unstable aircraft system illustrates the effectiveness of the proposed new method.

Keywords: Filter design from data, fault estimation, system inversion.

1. INTRODUCTION

Model-based fault diagnosis techniques for linear dynamic systems have been well established during the past two decades (Chen and Patton, 1999; Ding, 2013). However, an explicit and accurate system model is often unknown in practice. In such situations, a conventional approach first identifies the plant model from system input/output (I/O) data, and then adopts various model-based fault diagnosis methods (Simani et al., 2003; Patwardhan and Shah, 2005; Manuja et al., 2009). Without explicitly identifying a plant model, recent research efforts investigate data-driven approaches to directly construct a fault diagnosis system for additive sensor or actuator faults by utilizing the link between system identification and the model-based fault diagnosis methods (Qin, 2012; Ding, 2014a). These recent direct data-driven approaches simplify the design procedure by skipping the realization of an explicit plant model, while at the same time allow developing systematic methods to address the same fault diagnosis performance criteria as the existing model-based approaches.

Existing methods for data-driven fault detection and isolation construct residual generators with either the parity vector/matrix (Ding et al., 2009) or the Markov parameters (or impulse response parameters) (Dong et al., 2012a,b) that can be identified from data. Compared to generating residual signals sensitive to faults, fault estimation is much more involved. Dunia and Qin (1998) and

Qin (2012) proposed to reconstruct faults by minimizing the squared reconstructed prediction error in the residual subspace of a latent variable model. However, fault reconstructability of this approach was limited by the dimension of the residual subspace, especially when using a dynamic latent variable model (Dunia and Qin, 1998). Chapter 10 of Ding (2014a) first constructed a diagnostic observer realized with the identified parity vector/matrix, and then addressed faults as augmented state variables. This augmented observer scheme, however, imposed certain limitations on how fault signals vary with time. In contrast, Dong and Verhaegen (2012) constructed a system-inversion-based fault estimator using the identified Markov parameters, without any assumptions on the dynamics of fault signals. The drawback of this system-inversion-based method is that it cannot be applied to sensor faults in an unstable open-loop plant because its underlying system inverse used for the data-driven design is unstable in this case. In order to address this above drawback, we have recently proposed a receding horizon fault estimator by following a least-squares (LS) formulation of the fault estimation problem, and developed an H_2/H_∞ optimal design to compensate for identification errors of the Markov parameters (Wan et al., 2014b). This receding horizon method processes a batch of measurements at each time instant, thus may require increased computational effort.

This paper focuses on the direct data-driven design problem for a sensor fault estimation filter. This problem is challenging, especially when omitting the step of building an explicit state-space plant model in data-driven design, because the inverse of the underlying plant dynamics is

^{*} The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7-RECONFIGURE/2007–2013) under grant agreement No. 314544.

required and needs to be stable. In order to pave the way for the data-driven design, we first construct a system-inversion-based fault estimation filter based on the dynamics of the one-step-ahead predicted residual signal. The system inverse is divided into two parts: the open-loop left inverse, and the feedback from the residual reconstruction error to stabilize the inverse dynamics. This turns out to be stabilizable as long as the subsystem from faults to the outputs has no unstable invariant zeros. Our data-driven design method is obtained by parameterizing the above two parts of the inverse dynamics with the predictor Markov parameters identified from data.

Compared to the model-based approach based on an identified plant model, our direct data-driven design improves estimation performance by avoiding the propagation of model reduction errors originating from identification of the state-space plant model into the designed filter. Moreover, our proposed new method allows additional design freedom to stabilize the obtained filter under the same stabilizability condition as the existing model-based system inversion. This important additional property enables its application to sensor faults in unstable plants, where existing data-driven filter designs (Dong and Verhaegen, 2012) could not be applied so far due to the lack of such stability guarantees (even after stabilizing the closed-loop system). We also analyze the relationship between our novel data-driven design described in this paper, and our recently proposed moving horizon fault estimation method in Wan et al. (2014b). The presented new data-driven filter achieves better computational efficiency at the cost of minor performance loss and a more strict condition required for unbiasedness. The above significant advances to the state-of-the-art in data-driven fault estimation are illustrated via a numerical simulation example of an unstable aircraft system.

2. PRELIMINARIES AND PROBLEM FORMULATION

2.1 Notations

For a matrix X , X^- represents the left inverse satisfying $X^-X = I$, and $X^{(1)}$ represents the generalized inverse satisfying

$$XX^{(1)}X = X. \quad (1)$$

The i^{th} column of X is denoted by $X^{[i]}$. For the state-space model (A, B, C, D) or the sequence of Markov parameters H_0, H_1, \dots, H_{L-1} , let \mathcal{O}_L and \mathcal{T}_L denote the extended observability matrix with L block elements and the lower triangular block-Toeplitz matrix with L block columns and rows, respectively, i.e.,

$$\mathcal{O}_L(A, C) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{L-1} \end{bmatrix}, \mathcal{T}_L(\{H_i\}) = \begin{bmatrix} H_0 & 0 & \dots & 0 \\ H_1 & H_0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ H_{L-1} & H_{L-2} & \dots & H_0 \end{bmatrix}, \quad (2)$$

$$\text{or } \mathcal{T}_L(A, B, C, D) = \begin{bmatrix} D & 0 & \dots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ CA^{L-2}B & CA^{L-3}B & \dots & D \end{bmatrix}. \quad (3)$$

\mathbb{E} represents the mathematical expectation.

2.2 System description

We consider linear discrete-time systems governed by the following state-space model:

$$\begin{aligned} \xi(k+1) &= A\xi(k) + Bu(k) + Ef(k) + Fw(k) \\ y(k) &= C\xi(k) + Du(k) + Gf(k) + v(k). \end{aligned} \quad (4)$$

Here $\xi(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^{n_y}$, and $u(k) \in \mathbb{R}^{n_u}$ represent the state, the output measurement, and the known control input at time instant k , respectively. The process noise $w(k) \in \mathbb{R}^{n_w}$ and the measurement noise $v(k) \in \mathbb{R}^{n_v}$ are white zero-mean Gaussian, with covariance matrices $\mathbb{E}(w(k)w^T(k)) = Q$, $\mathbb{E}(v(k)v^T(k)) = R$, $\mathbb{E}(w(k)v^T(k)) = 0$. $f(k) \in \mathbb{R}^{n_f}$ is the unknown fault signal to be estimated. A, B, C, D, E, F, G are constant real matrices, with bounded norms and appropriate dimensions.

The following assumption is standard in Kalman filtering (Kailath et al., 2000) and subspace identification (Chiuso, 2007a,b):

Assumption 1. The pair (A, C) is assumed detectable; and there are no uncontrollable modes of $(A, FQ^{\frac{1}{2}})$ on the unite circle, where $Q^{\frac{1}{2}} \cdot (Q^{\frac{1}{2}})^T = Q$ is the covariance matrix of $w(k)$.

We consider additive sensor faults in this paper, i.e.,

$$E = 0_{n_x \times 1}, \quad G = I^{[j]} \quad (5)$$

for faults of the j^{th} sensor, with $X^{[j]}$ representing the j^{th} column of a matrix X . As in Dong and Verhaegen (2012), we adopt the following common assumption for sensor faults:

Assumption 2. $\text{rank}(G) = n_f$.

For data-driven design without knowing the system matrices in (4), it should be noted that in practice data collected under faulty conditions may be seldomly available, or if recorded then without a reliable fault description (Ding, 2014b). Hence we make the assumption as below:

Assumption 3. Only I/O data collected under the fault-free condition are used in our data-driven design.

No assumption is made in this paper about how the fault signals $f(k)$ evolve with time.

2.3 Problem formulation

With Assumption 1, the system (4) admits the innovation form given by

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + Ef(k) + Ke(k) \\ y(k) &= Cx(k) + Du(k) + Gf(k) + e(k). \end{aligned} \quad (6)$$

where K is the steady-state Kalman gain, $\{e(k)\}$ is the zero-mean innovation process with the covariance matrix Σ_e . Then $e(k)$ can be eliminated from the first equation of (6) to yield the one-step-ahead predictor form

$$\begin{aligned} x(k+1) &= \Phi x(k) + \tilde{B}u(k) + \tilde{E}f(k) + Ky(k) \\ y(k) &= Cx(k) + Du(k) + Gf(k) + e(k), \end{aligned} \quad (7)$$

with $\Phi = A - KC$, $\tilde{B} = B - KD$, and $\tilde{E} = E - KG$. The sensor fault direction matrices \tilde{E} and G in the predictor form (7) can be explicitly written as

$$\tilde{E} = -K^{[j]}, \quad G = I^{[j]} \quad (8)$$

for faults of the j^{th} sensor according to (5).

Denote the predictor Markov parameters by

$$H_i^u = \begin{cases} D & i = 0 \\ C\Phi^{i-1}\tilde{B} & i > 0 \end{cases}, H_i^y = \begin{cases} 0 & i = 0 \\ C\Phi^{i-1}K & i > 0 \end{cases}, \quad (9)$$

$$H_i^f = \begin{cases} G & i = 0 \\ C\Phi^{i-1}\tilde{E} & i > 0 \end{cases}.$$

With access only to the closed-loop data collected under the fault-free condition, the conventional model-based approach needs to identify the state-space model (7) for the fault estimation filter design. Such an identification algorithm follows three steps: (i) consistent LS estimation of the sequence of Markov parameters related to the fault-free subsystem $(\Phi, [\tilde{B} \ K], C, [D \ 0])$, i.e.,

$$\Xi = [H_p^u \ H_p^y \ \dots \ H_1^u \ H_1^y \ H_0^u]; \quad (10)$$

(ii) state-space realization of the fault-free subsystem $(\Phi, [\tilde{B} \ K], C, [D \ 0])$; (iii) construction of the fault direction matrices \tilde{E} and G according to (8). The first two identification steps above can follow the predictor-based subspace identification (PBSID) method in Chiuso (2007a,b). With the identified state-space model, existing model-based design approaches can be adopted. A disadvantage of the above design procedure is that the model reduction errors introduced in the state-space realization step would propagate into the fault estimation filter and might result in large fault estimation errors.

In order to avoid propagating the above model reduction errors into the designed filter, this paper aims to directly construct a stable sensor fault estimation filter with the Markov parameters Ξ identified from data.

3. SYSTEM-INVERSION-BASED FAULT ESTIMATION FILTER USING PREDICTOR FORM

As the foundation for our data-driven design, we construct a system-inversion-based fault estimation filter in this section using the state-space model of the predictor (7).

3.1 Open/Closed-loop left inverse

From (7), we construct a residual generator as follows:

$$\begin{aligned} \hat{x}(k+1) &= \Phi\hat{x}(k) + \tilde{B}u(k) + Ky(k) \\ r(k) &= y(k) - C\hat{x}(k) - Du(k), \end{aligned} \quad (11)$$

whose residual dynamics is

$$\tilde{x}(k+1) = \Phi\tilde{x}(k) + \tilde{E}f(k) \quad (12a)$$

$$r(k) = C\tilde{x}(k) + Gf(k) + e(k), \quad (12b)$$

with $\tilde{x}(k) = x(k) - \hat{x}(k)$.

By multiplying (12b) with G^- , it follows that $f(k)$ can be reconstructed as

$$f(k) = G^- (r(k) - C\tilde{x}(k) - e(k)).$$

Substituting the above equation into (12a) then yields the following left inverted system:

$$\tilde{x}(k+1) = \Phi_1\tilde{x}(k) + B_1(r(k) - e(k)) \quad (13a)$$

$$f(k) = C_1\tilde{x}(k) + D_1(r(k) - e(k)) \quad (13b)$$

with

$$\Phi_1 = \Phi - \tilde{E}G^-C, \quad B_1 = \tilde{E}G^-, \quad (14)$$

$$C_1 = -G^-C, \quad D_1 = G^-. \quad (15)$$

Since the innovation signal $e(k)$ and the initial state $\tilde{x}(0)$ are unknown, we construct the following system based on

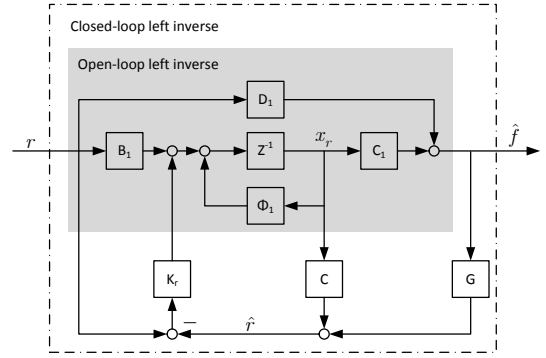


Fig. 1. Structure of the closed-loop left inverse

the inverted system (13) by ignoring $e(k)$ and replacing $\tilde{x}(k)$ and $f(k)$ with the state estimate $x_r(k)$ and the fault estimate $\hat{f}(k)$:

$$x_r(k+1) = \Phi_1 x_r(k) + B_1 r(k) \quad (16a)$$

$$\hat{f}(k) = C_1 x_r(k) + D_1 r(k). \quad (16b)$$

It is desired that the left inverse (16) is stable such that, starting from any arbitrary estimate $x_r(0)$ of the initial state $\tilde{x}(0)$, unbiasedness of the estimates $x_r(k)$ and $\hat{f}(k)$ can be achieved asymptotically. However, it is not guaranteed that the left inverse (16) is stable.

Next, we stabilize the inverted system (16) by feeding the residual reconstruction error back into (16a). Based on the state estimate $x_r(k)$ and the fault estimate $\hat{f}(k)$ in (16b), the residual signal $r(k)$ can be reconstructed as

$$\begin{aligned} \hat{r}(k) &= Cx_r(k) + G(C_1 x_r(k) + D_1 r(k)) \\ &= C_2 x_r(k) + D_2 r(k) \end{aligned} \quad (17)$$

according to (12b), with

$$C_2 = (I - GG^-)C, \quad D_2 = GG^-. \quad (18)$$

Then we construct the following closed-loop left inverse by feeding the residual reconstruction error $r(k) - \hat{r}(k)$ back into the open-loop left inverse (16):

$$\begin{aligned} x_r(k+1) &= \Phi_1 x_r(k) + B_1 r(k) + K_r (r(k) - \hat{r}(k)) \\ &= \Phi_2 x_r(k) + B_2 r(k) \end{aligned} \quad (19)$$

with

$$\Phi_2 = \Phi_1 - K_r C_2, \quad B_2 = B_1 + K_r (I - D_2). \quad (20)$$

Here “open-loop” and “closed-loop” refer to the absence/presence of the feedback from the residual reconstruction error in the two inverted systems (16) and (19). The structure of the closed-loop left inverse is illustrated in Figure 1.

It is worth noting that the data-driven filter design in Dong and Verhaegen (2012) considered only the open-loop left inverse whose stability is not guaranteed. It should also be pointed out that similarly to the simultaneous state and input estimation filter proposed in Gillijns (2007), the closed-loop left inverse (19) produces both the state estimate $x_r(k)$ and the fault estimate $\hat{f}(k)$. However, the proposed closed-loop inverse (19) has a more structured formulation, i.e., the combination of the open-loop left inverse and the feedback from the residual reconstruction error. Such a formulation enables our data-driven design in Section 4.

3.2 Stabilizability and unbiasedness

By defining $\tilde{x}_r(k) = \tilde{x}(k) - x_r(k)$ and $\tilde{f}(k) = f(k) - \hat{f}(k)$, we can obtain the dynamics of the fault estimation error as

$$\begin{aligned}\tilde{x}_r(k+1) &= \Phi_2 \tilde{x}_r(k) - B_2 e(k) \\ &= (\Phi_1 - K_r C_2) \tilde{x}_r(k) - B_2 e(k)\end{aligned}\quad (21)$$

$$\tilde{f}(k) = C_1 \tilde{x}_r(k) - D_1 e(k)$$

according to (12), (14)-(15), and (19)-(20). Therefore, if (Φ_1, C_2) is stabilizable, there exists a stabilizing gain K_r in (21), such that the obtained fault estimates are asymptotically unbiased, i.e., $\lim_{k \rightarrow \infty} \mathbb{E}(\tilde{f}(k)) = 0$.

Theorem 1. (Φ_1, C_2) is stabilizable if the fault subsystem (Φ, \tilde{E}, C, G) has no unstable invariant zeros.

The proof is given in Appendix A.

3.3 Fault estimation filter

So far we have constructed the closed-loop left inverse (19) to estimate faults from the residual signal generated by (11). By cascading the residual generator (11) and the closed-loop left inverse (19), we obtain the following fault estimation filter that produces fault estimates from the system inputs and outputs:

$$\begin{aligned}\begin{bmatrix} x_r(k+1) \\ \hat{x}(k+1) \end{bmatrix} &= \begin{bmatrix} \Phi_2 & -B_2 C \\ 0 & \Phi \end{bmatrix} \begin{bmatrix} x_r(k) \\ \hat{x}(k) \end{bmatrix} + \begin{bmatrix} -B_2 D & B_2 \\ \tilde{B} & K \end{bmatrix} z(k) \\ \hat{f}(k) &= [C_1 \quad -D_1 C] \begin{bmatrix} x_r(k) \\ \hat{x}(k) \end{bmatrix} + [-D_1 D \quad D_1] z(k)\end{aligned}\quad (22)$$

where

$$z(k) = [u^T(k) \quad y^T(k)]^T. \quad (23)$$

With $x_f(k) = x_r(k) + \hat{x}(k)$, the above fault estimation filter (22) can be reduced as follows by eliminating the unobservable modes:

$$\begin{aligned}x_f(k+1) &= (\Phi_1 - K_r C_2) x_f(k) + (B_f - K_r D_{f,2}) u(k) \\ &\quad + (K_f + K_r G_{f,2}) y(k) \\ \hat{f}(k) &= C_1 x_f(k) + D_{f,1} u(k) + D_1 y(k),\end{aligned}\quad (24)$$

where Φ_1 , C_1 , D_1 and C_2 are defined in (14), (15), and (18), respectively, and

$$B_f = \tilde{B} - \tilde{E} G^- D, \quad D_{f,2} = (I - G G^-) D, \quad (25a)$$

$$K_f = K + \tilde{E} G^-, \quad G_{f,2} = (I - G G^-), \quad (25b)$$

$$D_{f,1} = -G^- D. \quad (25c)$$

4. FAULT ESTIMATION FILTER DESIGN USING MARKOV PARAMETERS

In this section, we propose our Markov-parameter based design of the fault estimation filter (24) by exploiting its extended form over a time window.

4.1 Extended form of the fault estimation filter

With $k_0 = k - L + 1$, define stacked data vectors in a sliding window $[k_0, k]$ as $\mathbf{z}_{k,L}$, $\mathbf{r}_{k,L}$, $\mathbf{f}_{k,L}$, and $\mathbf{e}_{k,L}$, respectively for the signals z , r , f , and e , e.g.,

$$\mathbf{z}_{k,L} = [z^T(k_0) \cdots z^T(k)]^T. \quad (26)$$

Define \mathbf{T}_L^f and \mathbf{T}_L^z as the lower triangular block-Toeplitz matrices, i.e.,

$$\mathbf{T}_L^f = \mathcal{T}_L(\Phi, \tilde{E}, C, G), \quad \mathbf{T}_L^z = \mathcal{T}_L(\Phi, [\tilde{B} \ K], -C, [-D \ I]), \quad (27)$$

with \mathcal{O}_L and \mathcal{T}_L defined in (2) and (3). According to (11) and (12), the stacked residual signal $\mathbf{r}_{k,L}$ over the time window $[k_0, k]$ can be written in the extended form

$$\begin{aligned}\mathbf{r}_{k,L} &= \mathcal{O}_L(\Phi, -C) \cdot \hat{x}(k_0) + \mathbf{T}_L^z \mathbf{z}_{k,L} \\ &= \mathcal{O}_L(\Phi, C) \cdot \tilde{x}(k_0) + \mathbf{T}_L^f \mathbf{f}_{k,L} + \mathbf{e}_{k,L}\end{aligned}\quad (28)$$

where $\mathbf{z}_{k,L}$ is defined in (26), and $z(k)$ is defined in (23).

By assuming the initial state $x_r(k_0) = 0$ in the closed-loop left inverses (19), the stacked fault estimates $\hat{\mathbf{f}}_{k,L}$ over the time window $[k_0, k]$ can be written in the extended form

$$\hat{\mathbf{f}}_{k,L} = \mathcal{K}_L \cdot \mathbf{r}_{k,L} \quad \text{with} \quad \mathcal{K}_L = \mathcal{T}_L(\Phi_2, B_2, C_1, D_1). \quad (29)$$

With tedious but straightforward manipulations, the block-Toeplitz matrix \mathcal{K}_L defined in (29) can be rewritten as

$$\mathcal{K}_L = \mathcal{G}_L + \mathcal{M}_L (I - \mathbf{T}_L^f \mathcal{G}_L), \quad (30)$$

where

$$\mathcal{G}_L = \mathcal{T}_L(\Phi_1, B_1, C_1, D_1), \quad (31)$$

$$\mathcal{M}_L = \mathcal{T}_L(\Phi_1 - K_r C_2, K_r, C_1, 0). \quad (32)$$

By substituting (30) into (29), the extended form (30) can be intuitively explained:

- (i) $\mathcal{G}_L \mathbf{r}_{k,L}$ is the stacked fault estimates from the open-loop inverse (16);
- (ii) $\mathbf{T}_L^f \mathcal{G}_L \mathbf{r}_{k,L}$ is the stacked reconstructed residuals generated by (16a) and (17);
- (iii) \mathcal{M}_L represents the feedback dynamics from the stacked residual reconstruction errors $\mathbf{r}_{k,L} - \mathbf{T}_L^f \mathcal{G}_L \mathbf{r}_{k,L}$ to the stacked fault estimates.

By substituting (28) and (30) into (29), we obtain the following extended form of the fault estimation filter (24):

$$\hat{\mathbf{f}}_{k,L} = \mathcal{O}_L(\Phi_2, C_1) \cdot \hat{x}(k_0) + (\mathcal{R}_L + \mathcal{M}_L \mathcal{Q}_L) \cdot \mathbf{z}_{k,L}, \quad (33a)$$

$$= \mathcal{O}_L(\Phi_2, -C_1) \cdot \tilde{x}(k_0) + \mathbf{f}_{k,L} + \mathcal{K}_L \mathbf{e}_{k,L} \quad (33b)$$

with

$$\mathcal{R}_L = \mathcal{G}_L \cdot \mathbf{T}_L^z = \mathcal{T}_L(\Phi_1, [B_f \ K_f], C_1, [D_{f,1} \ D_1]),$$

$$\begin{aligned}\mathcal{Q}_L &= (I - \mathbf{T}_L^f \mathcal{G}_L) \cdot \mathbf{T}_L^z = \mathbf{T}_L^z - \mathbf{T}_L^f \mathcal{R}_L \\ &= \mathcal{T}_L(\Phi_1, [B_f \ K_f], -C_2, [-D_{f,2} \ G_{f,2}]).\end{aligned}\quad (34)$$

The derivation of \mathcal{R}_L in (34) is obtained by regarding $\mathcal{G}_L \cdot \mathbf{T}_L^z$ as cascading the system $(\Phi, [\tilde{B} \ K], -C, [-D \ I])$ related to \mathbf{T}_L^z and the system (Φ_1, B_1, C_1, D_1) related to \mathcal{G}_L . Similar methods apply to the derivation of \mathcal{Q}_L in (34).

It can be seen from (33b) that $\hat{\mathbf{f}}_{k,L}$ is a biased estimate of $\mathbf{f}_{k,L}$ due to the presence of unknown initial state $\tilde{x}(k_0)$. However, it follows from (33b) and the definition of $\mathcal{O}_L(\Phi_2, -C_1)$ in (2) that

$$\mathbb{E}(\hat{f}(k) - f(k)) = -C_1 \Phi_2^{L-1} \tilde{x}(k_0),$$

where $\hat{f}(k)$ and $f(k)$ are the last n_f entries of $\hat{\mathbf{f}}_{k,L}$ and $\mathbf{f}_{k,L}$, respectively. The above equation shows that $\hat{f}(k)$, extracted from $\hat{\mathbf{f}}_{k,L}$ in (33a), gives asymptotically unbiased fault estimation as L goes to infinity if Φ_2 is stabilized given the condition in Theorem 1.

4.2 Markov-parameter based design

In order to avoid propagating model reduction errors that originate from the identification of the state-space plant model into the designed filter, we now directly construct the sensor fault estimation filter (24) from data by utilizing its extended form (33a). The basic idea follows four steps:

- (i) Identify the Markov parameters Ξ in (10) using the plant I/O data (Chiuso, 2007a,b);
- (ii) Compute the Markov parameters of the system

$$\left(\Phi_1, [B_f \ K_f], \begin{bmatrix} C_1 \\ -C_2 \end{bmatrix}, \begin{bmatrix} D_{f,1} & D_1 \\ -D_{f,2} & G_{f,2} \end{bmatrix} \right) \quad (35)$$

which combines the dynamics related to \mathcal{R}_L and \mathcal{Q}_L in (34);

- (iii) Find a state-space realization of the system (35);
- (iv) Find a stabilizing gain L_2 for (Φ_1, C_2) , and then construct the fault estimation filter (24) with the identified system matrices in (35).

In Step (ii), it should be noted that Ξ in (10) includes only the Markov parameters $\{H_i^u\}$ and $\{H_i^y\}$ related to system inputs and outputs. According to (8) and (9), the Markov parameters $\{H_i^f\}$ related to the j^{th} sensor faults and $\{H_i^z\}$ related to \mathbf{T}_L^z in (27) can be obtained as

$$H_i^f = \begin{cases} I^{[j]} & i = 0 \\ -(H_i^y)^{[j]} & i > 0 \end{cases}, \quad H_i^z = \begin{cases} [-H_0^u \ I] & i = 0 \\ [-H_i^u \ -H_i^y] & i > 0 \end{cases}. \quad (36)$$

Let $\{G_i\}$, $\{R_i\}$ and $\{Q_i\}$ denote the Markov parameters that construct the block-Toeplitz matrices

$$\mathcal{G}_L = \mathcal{T}_L(\{G_i\}), \quad \mathcal{R}_L = \mathcal{T}_L(\{R_i\}), \quad \mathcal{Q}_L = \mathcal{T}_L(\{Q_i\})$$

with the definition of \mathcal{T}_L in (2). In order to ensure $\mathcal{G}_L \mathbf{T}_L^f = I$, the Markov parameters $\{G_i\}$ can be computed as

$$\begin{cases} G_0 = (H_0^f)^-, \\ G_i = -\sum_{j=1}^i G_{i-j} H_j^f G_0, \quad 1 \leq i \leq L-1. \end{cases} \quad (37)$$

According to the definition of \mathcal{R}_L in (34), its Markov parameters $\{R_i\}$ can be computed as the convolution of $\{G_i\}$ in (37) and $\{H_i^z\}$ in (36):

$$R_i = \sum_{j=0}^i G_{i-j} H_j^z, \quad i = 0, \dots, L-1.$$

Similarly, the Markov parameters $\{Q_i\}$ of \mathcal{Q}_L in (34) can be computed as

$$Q_i = H_i^z - \sum_{j=0}^i H_{i-j}^f R_j, \quad i = 0, \dots, L-1.$$

Since the system (35) combines the dynamics of \mathcal{R}_L and \mathcal{Q}_L in (34), the Markov parameters of the system (35) are $\{W_i\}$ with $W_i = \begin{bmatrix} R_i \\ Q_i \end{bmatrix}$.

In Step (iii), using the Markov parameters $\{W_i\}$ of the system (35), it is straightforward to obtain

$$\begin{bmatrix} \hat{D}_{f,1} & \hat{D}_1 \\ -\hat{D}_{f,2} & \hat{G}_{f,2} \end{bmatrix} = W_0 = \begin{bmatrix} R_0 \\ Q_0 \end{bmatrix}, \quad (38)$$

and formulate the block Hankel matrix

$$\mathcal{H}_W = \begin{bmatrix} W_1 & W_2 & \dots & W_m \\ W_2 & W_3 & \dots & W_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ W_l & W_{l+1} & \dots & W_{l+m-1} \end{bmatrix}, \quad (39)$$

which corresponds to the system (35). Then, compute the singular value decomposition (SVD) of \mathcal{H}_W , i.e.,

$$\mathcal{H}_W = [U_W \ U_W^\perp] \begin{bmatrix} \Sigma_W & 0 \\ 0 & \Sigma_W^\perp \end{bmatrix} \begin{bmatrix} V_W^T \\ (V_W^\perp)^T \end{bmatrix}.$$

In this above equation, the nonsingular and diagonal matrix Σ_W consists of the \hat{n} largest singular values of the block Hankel matrix \mathcal{H}_W , where \hat{n} is actually the selected order of the fault estimation filter (24). The order \hat{n} can be chosen by examining the gap among the singular values of \mathcal{H}_W , as in subspace identification methods (Chiuso, 2007a). Now, we can write

$$\hat{\mathcal{H}}_W = U_W \Sigma_W V_W^T. \quad (40)$$

From (40), the estimated controllability and observability matrices can be constructed as (Chiuso, 2007a)

$$\hat{\mathcal{C}}_W = \Sigma_W^{\frac{1}{2}} V_W^T, \quad \hat{\mathcal{O}}_W = U_W \Sigma_W^{\frac{1}{2}}. \quad (41)$$

Then the state-space matrices of the system (35) are computed as below:

$$[\hat{B}_f \ \hat{K}_f] = \text{the first } n_u + n_y \text{ columns of } \hat{\mathcal{C}}_W, \quad (42)$$

$$\begin{bmatrix} \hat{C}_1 \\ -\hat{C}_2 \end{bmatrix} = \text{the first } n_f + n_y \text{ rows of } \hat{\mathcal{O}}_W, \quad (43)$$

$$\hat{\Phi}_1 = \hat{\mathcal{C}}_{W,2} \hat{\mathcal{C}}_{W,1}^T \left(\hat{\mathcal{C}}_{W,1} \hat{\mathcal{C}}_{W,1}^T \right)^{-1}, \quad (44)$$

where $\hat{\mathcal{C}}_{W,1}$ and $\hat{\mathcal{C}}_{W,2}$ are the matrices consisting of the first and, respectively, the last $n_u(m-1)$ columns of $\hat{\mathcal{C}}_W$.

Finally, we follow Step (iv) by using the estimated state-space matrices in (38) and (42)-(44).

4.3 Comparisons and discussion

Fault reconstruction based on dynamic latent variable models has been widely adopted in statistical process monitoring (Qin, 2012). In these types of methods, fault estimates are obtained by minimizing the squared reconstructed prediction error in the residual subspace of a latent variable model. By analyzing its close link with subspace identification, Qin and Li (2001) showed that the dimension of the residual subspace is determined by the left null space of the observability matrix, thus limiting fault reconstructability for dynamic latent variable models.

Although the stabilizability condition in Theorem 1 is derived by using the predictor model (7), it can still be applied to the data-driven filter design without requiring the state-space model, because the invariant zeros of the underlying fault subsystem (Φ, \hat{E}, C, G) can be checked by using the identified Markov parameters $\{H_i^f\}$ (Yeung and Kwan, 1993; Fledderjohn et al., 2010). In this sense, the stabilizability condition of our proposed data-driven filter design is the same as in model-based filter design.

The additional design freedom for stabilization is not possible in the data-driven fault estimation filter design proposed by Dong and Verhaegen (2012), because it considered only the open-loop left inverse (16) and missed the

feedback part in the inverse dynamics. For the same reason, the data-driven filter in Dong and Verhaegen (2012) would be unstable in some situations, such as sensor faults of an unstable open-loop plant. In contrast, our data-driven design is based on the closed-loop left inverse (19) which guarantees the stability and unbiasedness of the constructed fault estimation filter under the condition given by Theorem 1.

The above discussion explains why sensor faults of an unstable open-loop plant cannot be tackled by applying the data-driven method in Dong and Verhaegen (2012) to the open-loop plant. It is worth noting that this difficulty cannot be solved by simply applying the same method to the stabilized closed-loop system. The reason is that the sensor faults affect not only output equations but also the closed-loop dynamics, hence the first equation in (36) is no longer valid for the Markov parameters $\{H_i^f\}$ of the closed-loop fault subsystem. In fact, with only fault-free I/O data as stated in Assumption 3, there is no simple way to derive $\{H_i^f\}$ from the identified Markov parameters $\{H_i^u, H_i^y\}$ of a closed-loop system (Wan and Ye, 2012). In order to use the closed-loop system for data-driven sensor fault estimation, Section V.B of Dong and Verhaegen (2012) proposed to use a special control law such that the sensor faults did not affect the closed-loop dynamics, which is not always possible in practice.

It is also interesting to investigate the computational complexity of our proposed fault estimation filter (24), which turns out to be $O((n_f + n_x)(n_x + n_u + n_y))$ at each time instant. In order to put this in perspective, we compare it with our recently proposed moving horizon fault estimator in Wan et al. (2014b), where choosing a certain horizon length L , the computations at each time instant involve multiplying a matrix of size $n_f \times L(n_u + n_y)$ with $\mathbf{z}_{k,L}$ defined in (26). This leads to a computational complexity of $O(Ln_f(n_u + n_y))$. Therefore, by directly identifying the fault estimation filter (24), our proposed new method would be more efficient in terms of computational complexity, since L needs to be sufficiently large in order to achieve asymptotic unbiasedness in Wan et al. (2014b).

We show in Appendix B that our recently proposed moving horizon fault estimator in Wan et al. (2014b) can be equivalently rewritten as

$$\hat{\mathbf{f}}'_{k,L} = \left[\mathcal{G}'_L + \mathcal{M}'_L \left(I - \mathbf{T}_L^f \mathcal{G}'_L \right) \right] \cdot \mathbf{r}_{k,L} \quad (45)$$

which is in the same form of (29) and (30) except that \mathcal{G}_L and \mathcal{M}_L in (30) have the block-Toeplitz structure defined in (32) while \mathcal{G}'_L and \mathcal{M}'_L in (45) are dense matrices according to (B.4)-(B.6). Compared to the moving horizon fault estimator in Wan et al. (2014b), the block-Toeplitz structure of \mathcal{G}_L and \mathcal{M}_L in our proposed new design affects two aspects:

- (i) It has been proved in Wan et al. (2014a) that with dense matrices \mathcal{G}'_L and \mathcal{M}'_L in (45), the LS moving horizon fault estimation used in Wan et al. (2014b) gives minimal variance over all linear unbiased estimators. In contrast, the block-Toeplitz structure of \mathcal{G}_L and \mathcal{M}_L in our proposed new method leads to less design freedom and minor performance loss.

- (ii) In order to ensure asymptotically unbiased estimation, the moving horizon fault estimator in Wan et al. (2014b) requires that the fault subsystem has no unstable transmission zeros, while a more strict condition is needed for our proposed new design, i.e., not only the transmission zeros but all the invariant zeros of the fault subsystem have to be stable.

5. SIMULATION STUDIES

Consider the linearized continuous-time VTOL (vertical takeoff and landing) aircraft model used in Dong and Verhaegen (2012). The model has four states, namely horizontal velocity, vertical velocity, pitch rate, and pitch angle. The two inputs are collective pitch control and longitudinal cyclic pitch control. With a sampling rate of 0.5 seconds, the discrete-time model (4) is obtained, with $D = 0$ and $F = I_4$. The process and measurement noise, $w(k)$, $v(k)$, are zero mean white, respectively with a covariance of $Q_w = 10^{-4} \cdot I_4$ and $Q_v = 0.01 \cdot I_2$. Since the open-loop plant is unstable, an empirical stabilizing output feedback controller is used, i.e.,

$$u(k) = - \begin{bmatrix} 0 & 0 & -0.5 & 0 \\ 0 & 0 & -0.1 & -0.1 \end{bmatrix} \cdot y(k) + \eta(k), \quad (46)$$

where $\eta(k)$ is the reference signal. All the parameters of the plant and the controller are the same as those in Dong and Verhaegen (2012).

In the identification experiment, the reference signal $\eta(k)$ is zero-mean white noise with the covariance of $\text{diag}(1, 1)$, which ensures persistent excitation. We collect $N = 1000$ data samples from the identification experiment. In the identification algorithm, the past horizon is selected as $p = 100$.

The simulated sensor fault signals affect the first two sensors, i.e., $E = 0_{4 \times 2}$, $G = \begin{bmatrix} 1000 & 1 \\ 0 & 100 \end{bmatrix}^T$,

$$f(k) = \begin{cases} \begin{bmatrix} 0 & 0 \end{bmatrix}^T, & 0 \leq k \leq 50, \\ \begin{bmatrix} \sin(0.1\pi k) & 1 \end{bmatrix}^T, & k > 50. \end{cases}$$

When simulating sensor faults, the reference signal $\eta(k)$ in the control law (46) is set to be zero.

We will compare the following methods for fault estimation filter design:

- Alg0: design based on the accurate predictor model (7);
- Alg1: design based on the state-space model of the predictor (7) identified from data;
- Alg2: our proposed new method in Section 4.2;
- Alg3: our recently proposed moving horizon fault estimator constructed from the predictor Markov parameters Ξ identified from data (Wan et al., 2014b).

In the state-space realization from Hankel matrices, the orders of the plant model in Alg1 and the filter in Alg2 are both selected as 4, the same as the underlying plant. For Alg2 and Alg3, the order of the Markov parameters of the system (35) is $L = 100$, and the number of block rows and columns of the block Hankel matrix \mathcal{H}_W in (39) is $l = m = 20$. The poles of the first three fault estimation filters are placed at the same location, i.e., $[0.948, 0.532, 0.225, 0.141]^T$.

The method in (Dong and Verhaegen, 2012) cannot be directly applied to the above sensor fault scenario due to the reasons explained in Section 4.3, thus it is not implemented here. In contrast, all the three fault estimation filters above are based on the closed-loop left inverse (19), and their stability is guaranteed since the condition in Theorem 1 is satisfied in this simulated scenario.

The distributions of fault estimation errors are shown in Figure 2. Because of the noisy identification data and the model reduction errors, the three data-driven designs, Alg1, Alg2, and Alg3, all give larger estimation error covariances than Alg0 based on the accurate plant model. Figure 2 also clearly shows that our proposed Alg2 achieves better estimation performance than Alg1. This is because Alg2 is not subject to model reduction errors before realizing the state-space matrices in (35) as part of the fault estimation filter (24), while model reduction errors are introduced instead by Alg1 in the identified plant model, and propagated into larger uncertainties of the fault estimation filter. As explained at the end of Section 4.3, Alg2 is much faster than Alg3 at the cost of minor performance loss.

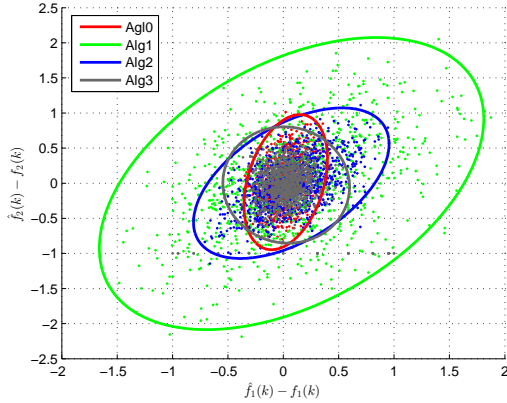


Fig. 2. Distribution of fault estimation errors. Dots: 1000 estimation errors by different fault estimation methods. Ellipses: the 3σ -contour of the approximated two-dimensional Gaussian distribution of the 1000 estimation errors, i.e., the contour at $[\hat{f}(k) - f(k)]^T \text{cov}^{-1}(\hat{f}(k)) [\hat{f}(k) - f(k)] = 3$.

6. CONCLUSIONS

A novel direct data-driven design method has been proposed for sensor fault estimation filters by parameterizing the system-inversion-based fault estimation filter with Markov parameters. The proposed approach simplifies the design procedure by omitting the step of identifying the state-space plant model, and improves estimation performance by avoiding the propagation of some model reduction errors into the designed filter. Moreover, it allows additional design freedom to stabilize the designed filter under the same stabilizability condition as model-based system inversion, thus can be applied to sensor faults in an unstable plants. Detailed analysis has been given to explain why sensor faults in an unstable plant cannot be tackled by applying existing state-of-the-art data-driven methods with predictor Markov parameters of either the unstable open-loop plant or the stabilized closed-loop

system. A numerical simulation example illustrates the effectiveness of our method applied to sensor faults of an unstable aircraft system, and the advantage of the direct data-driven design. Future work will focus on how to enhance robustness against both identification errors of Markov parameters and model reduction errors in realizing the state-space form of the filter.

REFERENCES

- Chen, J. and Patton, R. (1999). *Robust Model-Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic, Norwell, MA.
- Chiuso, A. (2007a). On the relation between CCA and predictor-based subspace identification. *IEEE Transactions on Automatic Control*, 52, 1795–1812.
- Chiuso, A. (2007b). The role of vector autoregressive modeling in predictor based subspace identification. *Automatica*, 43, 1034–1048.
- Ding, S.X. (2013). *Model-Based Fault Diagnosis Techniques: Design Scheme, Algorithms, and Tools*. Springer-Verlag, London, 2 edition.
- Ding, S.X. (2014a). *Data-Driven Design of Fault Diagnosis and Fault-Tolerant Control Systems*. Springer-Verlag, London.
- Ding, S.X. (2014b). Data-driven design of monitoring and diagnosis systems for dynamic processes: a review of subspace technique based schemes and some recent results. *Journal of Process Control*, 24, 431–449.
- Ding, S.X., Zhang, P., Naik, A., Ding, E., and Huang, B. (2009). Subspace method aided data-driven design of fault detection and isolation systems. *Journal of Process Control*, 19, 1496–1510.
- Dong, J. and Verhaegen, M. (2012). Identification of fault estimation filter from I/O data for systems with stable inversion. *IEEE Transactions on Automatic Control*, 57, 1347–1361.
- Dong, J., Verhaegen, M., and Gustafsson, F. (2012a). Robust fault detection with statistical uncertainty in identified parameters. *IEEE Transactions on Signal Processing*, 60, 5064–5076.
- Dong, J., Verhaegen, M., and Gustafsson, F. (2012b). Robust fault isolation with statistical uncertainty in identified parameters. *IEEE Transactions on Signal Processing*, 60, 5556–5561.
- Dunja, R. and Qin, S.J. (1998). Subspace approach to multidimensional fault identification and reconstruction. *AIChE Journal*, 44, 1813–1831.
- Fledderjohn, M.S., Holzel, M.S., Morozov, A.V., and Bernstein, D.S. (2010). On the accuracy of least squares algorithms for estimating zeros. In *Proceedings of 2010 American Control Conference*, 3729–3734. Baltimore, MD, USA.
- Gillijns, S. (2007). *Kalman Filtering Techniques for System Inversion and Data Assimilation*. Ph.D. thesis, Katholieke University Leuven.
- Kailath, T., Sayed, A., and Hassibi, B. (2000). *Linear Estimation*. Prentice-Hall, Englewood Cliffs, NJ.
- Manuja, S., Narasimhan, S., and Patwardhan, S.C. (2009). Unknown input modeling and robust fault diagnosis using black box observers. *Journal of Process Control*, 19, 25–37.
- Patwardhan, S.C. and Shah, S.L. (2005). From data to diagnosis and control using generalized orthonormal

basis filters. Part I: development of state observers. *Journal of Process Control*, 15, 819–835.

Qin, S.J. (2012). Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control*, 36, 220–234.

Qin, S.J. and Li, W. (2001). Detection and identification of faulty sensors in dynamic processes. *AIChE Journal*, 47, 1581–1593.

Simani, S., Fantuzzi, S., and Patton, R. (2003). *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. Springer-Verlag, London.

Wan, Y., Keviczky, T., and Verhaegen, M. (2014a). Moving horizon least-squares input estimation for linear discrete-time stochastic systems. In *Proc. IFAC World Congress*, 3483–3488. Cape Town, South Africa.

Wan, Y., Keviczky, T., Verhaegen, M., and Gustafsson, F. (2014b). Data-driven robust receding horizon fault estimation. Submitted to *Automatica*.

Wan, Y. and Ye, H. (2012). Data-driven diagnosis of sensor precision degradation in the presence of control. *Journal of Process Control*, 22, 26–40.

Yeung, K.S. and Kwan, C.M. (1993). System zeros determination from an unreduced matrix fraction description. *IEEE Transactions on Automatic Control*, 38, 1695–1697.

Appendix A. PROOF OF THEOREM 1

In order to prove (Φ_1, C_2) is stabilizable, we need to show that (Φ_1, C_2) has no unstable unobservable modes, i.e.,

$$\text{rank} \left(\begin{bmatrix} \Phi_1 - \lambda I \\ C_2 \end{bmatrix} \right) = n \text{ for } |\lambda| \geq 1. \quad (\text{A.1})$$

By following (14) and (18), it can be derived that

$$\begin{bmatrix} \Phi_1 - \lambda I & \tilde{E} \\ C_2 & G \end{bmatrix} = \begin{bmatrix} \Phi & -\lambda I & \tilde{E} \\ C & G \end{bmatrix} \cdot \begin{bmatrix} I & 0 \\ -G^- C & I \end{bmatrix}. \quad (\text{A.2})$$

With Assumption 2, if (Φ, \tilde{E}, C, G) has no unstable invariant zeros, it follows that

$$\text{rank} \left(\begin{bmatrix} \Phi_1 - \lambda I & \tilde{E} \\ C_2 & G \end{bmatrix} \right) = \text{rank} \left(\begin{bmatrix} \Phi & -\lambda I & \tilde{E} \\ C & G \end{bmatrix} \right) = n + n_f$$

for $|\lambda| \geq 1$, which implies (A.1).

Appendix B. PROOF OF (45)

With the definitions

$$\mathbf{O}_L = \mathcal{O}_L(\Phi, C), \quad \Psi_L = [\mathbf{O}_L \quad \mathbf{T}_L^f], \quad \mathbf{f}_{k,L}^x = \begin{bmatrix} \tilde{x}(k_0) \\ \mathbf{f}_{k,L} \end{bmatrix}, \quad (\text{B.1})$$

the LS fault estimation problem

$$\min_{\mathbf{f}_{k,L}^x} \|\mathbf{r}_{k,L} - \Psi_L \mathbf{f}_{k,L}^x\|_2^2 \quad (\text{B.2})$$

can be formulated based on the second equation of (28) (Wan et al., 2014a). The LS problem (B.2) has non-unique solutions because Ψ_L may not have full column rank. By using the generalized inverse defined in (1), one solution to the problem (B.2) is

$$\hat{\mathbf{f}}_{k,L}^x = (\Psi_L^T \Psi_L)^{(1)} \Psi_L^T \mathbf{r}_{k,L}. \quad (\text{B.3})$$

According to the definition of Ψ_L in (B.1) and Schur complements (Kailath et al., 2000), the estimate of $\mathbf{f}_{k,L}$, i.e., (45), can be extracted from (B.3), with

$$\mathcal{G}'_L = \left((\mathbf{T}_L^f)^T \mathbf{T}_L^f \right)^{-1} (\mathbf{T}_L^f)^T, \quad (\text{B.4})$$

$$\mathcal{M}'_L = \mathcal{G}'_L \mathbf{O}_L \Delta^{(1)} \mathbf{O}_L^T, \quad (\text{B.5})$$

$$\Delta = \mathbf{O}_L^T \mathbf{O}_L - \mathbf{O}_L^T \mathbf{T}_L^f \mathcal{G}'_L \mathbf{O}_L. \quad (\text{B.6})$$

Although $\hat{\mathbf{f}}_{k,L}'$ in (45) may be a biased estimate of $\mathbf{f}_{k,L}$ due to column rank deficiency of Ψ_L , its last n_f entries give an asymptotically unbiased estimate of $f(k)$ as L goes to infinity (Wan et al., 2014a).

Note that with the innovation covariance matrix estimated from data (Chiuso, 2007a,b), a weighted LS problem can be formulated for fault estimation (Wan et al., 2014a). For the sake of simplicity, an ordinary LS problem (B.2) is used here without loss of generality.